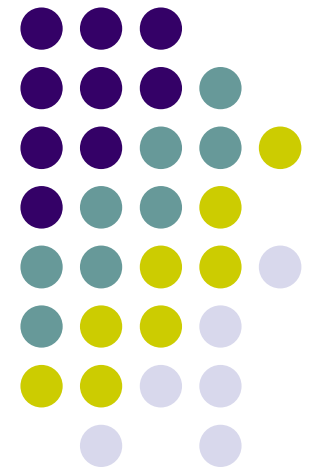


Information Engineering and Adaptive Data Systems

Todd A. King, Raymond J. Walker,
Steven P. Joy, D. Aaron Roberts,
James R. Thieman



**Paper SM11C-03
2006 Fall AGU Meeting
San Francisco, CA**

Strategic Objectives of the Heliophysics Data Environment



Enhance scientific research in heliophysics

Allow the scientist to focus on the question
rather than the nuances of data acquisition.

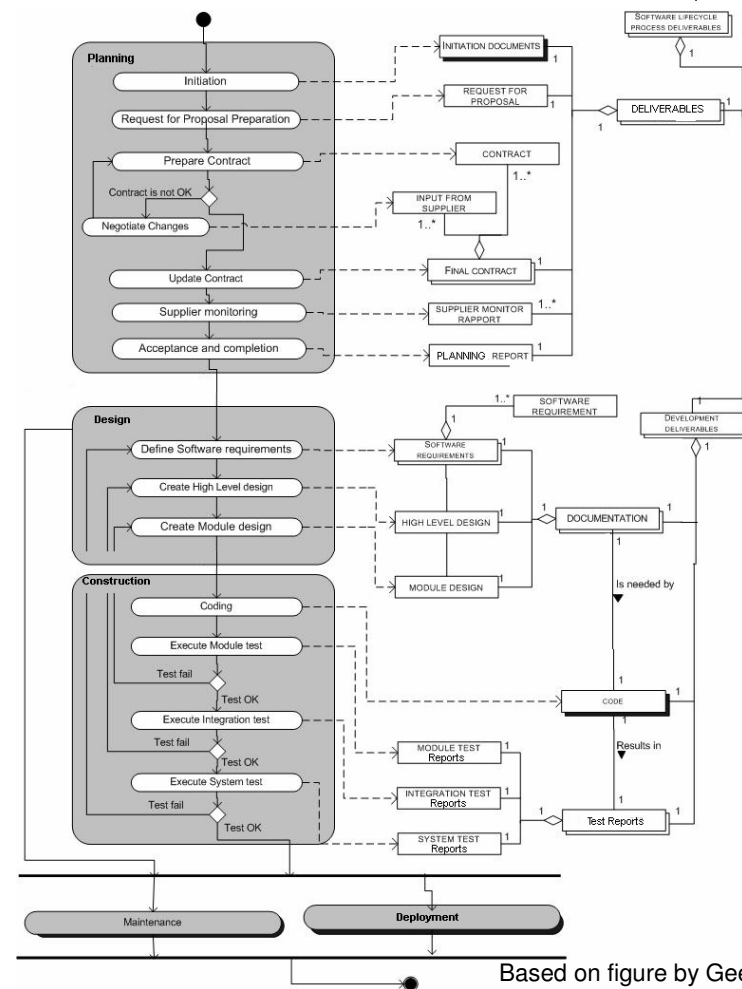


The Heliophysics Challenge

- Widely distributed holdings
 - Mission database
 - Individual investigators
 - Independent researchers
 - Resident archives
- Instruments are more sophisticated
 - Increasing data rates
- Legacy data is critically important
 - Long term influences (11 year solar cycle)
- Enhanced and value-added derivatives
 - Researchers are improving the body of knowledge
- Growing need to integrate
 - missions (projects)
 - spacecraft
 - instruments
 - discipline
- An international activity
 - Many independent efforts
 - Loosely coupled

Meeting the Challenge

- Build a system to manage and distribute information over an entire enterprise.
- This requires engineering-like methods.
 - **planning** (enterprise driven)
 - **analysis** (modeling)
 - **design** (documentation and diagrams)
 - **construction** (coding and testing)
 - **deployment** (distribution and installation)

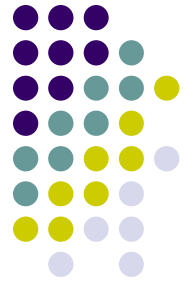


Based on figure by Geert Vermunt

Which Development Approach?



- There are a lot of development approaches:
 - Agile, Extreme Programming, Rapid Application Development (RAD), cowboy, iterative, Information Engineering Methodology (IEM), ...
- The difference between each of these approaches is
 - The emphasis of each phase
 - Extent and frequency of iterations (loops)
 - Scalability (to an enterprise)

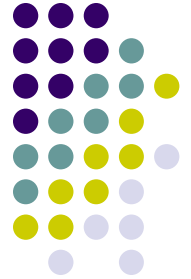


Information Engineering

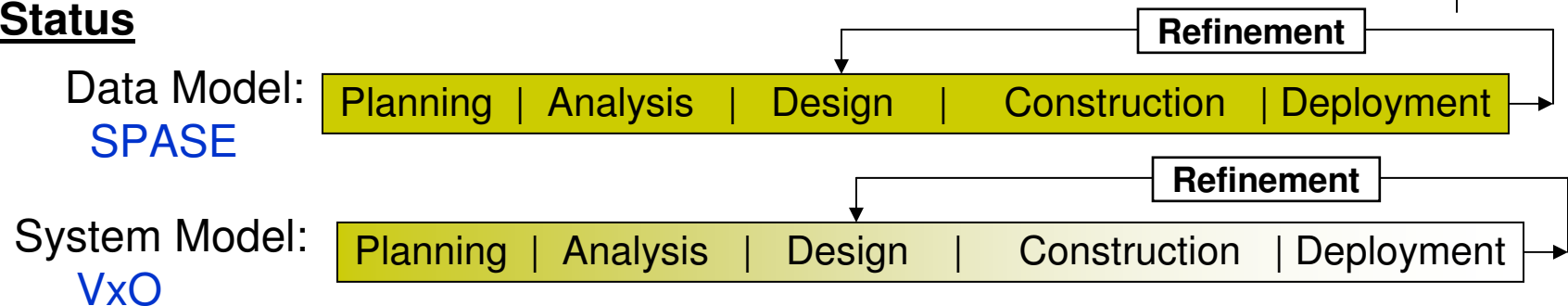
- Best choice: Information Engineering¹ Methodology (IEM)
 - Because it has been designed and refined to address the enterprise-wide challenges
 - It integrates data engineering, software engineering and system engineering into a single methodology
 - Project's vision is documented
 - Phases align with vision
 - Project divided into achievable units (scope and priority)
 - Multiple efforts with an integral purpose

1. Martin, James and Clive Finkelstein. Nov 1981. "Information Engineering", Technical Report (2 volumes), Savant Institute, Carnforth, Lancs, UK.

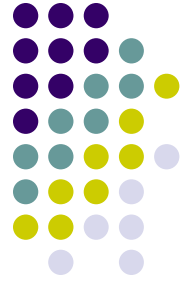
The State of the Heliophysics Data Environment



Status

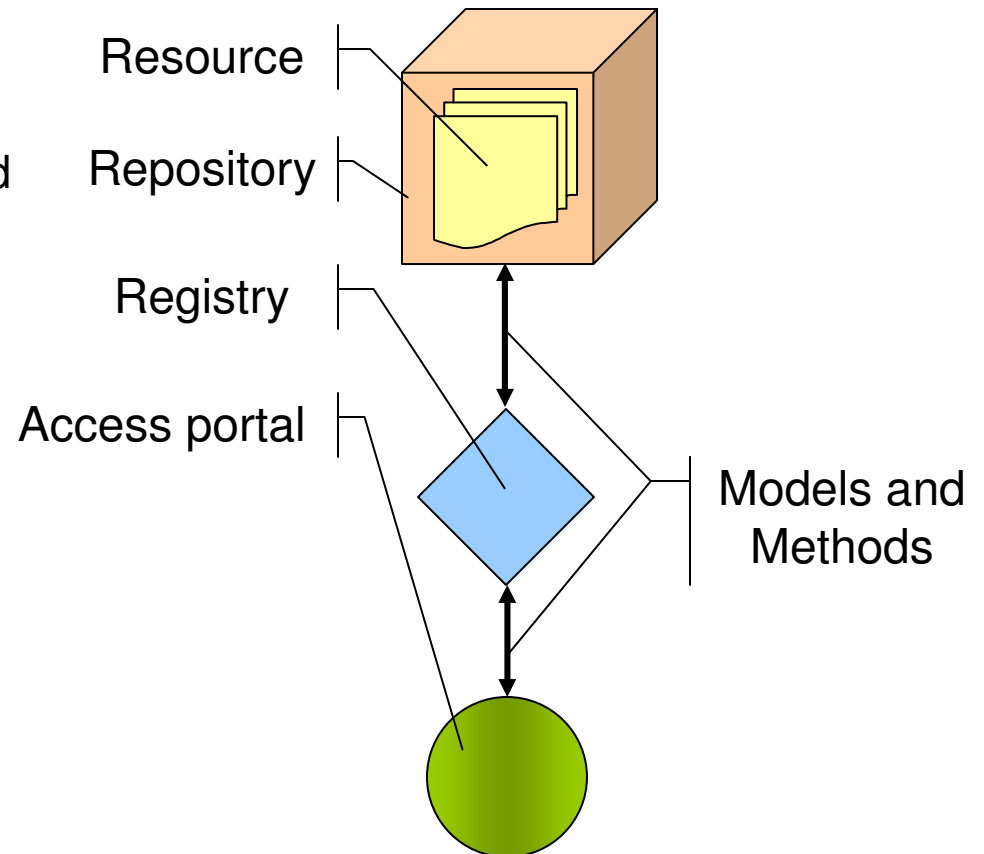


- Planning (enterprise driven)
 - The needs of the community
 - The VxO vision
- Analysis
- Design (documentation and diagrams)
 - SPASE system model
 - Individual VxO models
- Definition of tasks (smaller projects with an integral purpose)
 - SPASE (Modeling and standards)
 - Discipline oriented VxO
 - Data preparation tasks
- Construction
 - The VxOs are starting
- Deployment
 - In phases over the next year
- Refinement (maintenance)
 - SPASE data model is at version 1.1.0 and planning 1.2.0



Current System Model

- **Resource:** An object (document, data, etc.) or service available for use.
- **Repository:** A facility for storing and maintaining digital information in accessible form
- **Registry:** A collection point for metadata about resources.
- **Access Portal:** An interface to the registries and resources.
- **Data Model:** Describes in an abstract way how data is represented. This includes semantics (meaning of terms) and ontology (relationships).
- **Access Methods:** Mechanisms to search for, use and distribute resources.



The Heliophysics Virtual Observatories



- The VxO will:
 - Share a common data model (SPASE)
 - Exchange metadata in standard formats (XML)
 - Actively harvest (find out what's available)
 - Based on community contributions
 - Dynamically align with new programs
 - Layered framework (many active components)
 - Established as a service to be responsive to the user

We succeed if it is easier for the user to do science.

What type of system can do it all?



- Adaptive Data Systems
 - Data Systems which change in response to user expectations, provider needs and community input
- Adaptive Data Systems are:
 - Self-organizing
 - Self-managing
 - Proactive

Building an Adaptive Data System



- An adaptive data system can be achieved by:
 - Defining a framework for information exchange
 - protocols
 - formats
 - Allow for asynchronous share and harvest
 - Resources are made available (share)
 - Information about the resources is collected (harvest)
 - Develop autonomous agents



Conclusions

- Information Engineering methods can help guide the development of a heliophysics data system
 - Enterprise-wide activity
 - Must integrate all facets of engineering (data, system, software)
 - Must meet the strategic objective of aiding science
- Adaptive data systems fit the scenarios because:
 - Many different, simultaneous activities
 - Minimal management (cost effective)
 - Incremental evolution of the entire system
- If we use Information Engineering methodology as a roadmap what can we anticipate?
 - We can expect a network of VxOs that are focused on the needs of the scientist.



Thank You!

Backup Material





Abstract

- Information engineering is a rigorous architectural approach to developing and deploying applications within an enterprise. It progresses through natural stages of analysis, design and implementation with the focus on how information is organized to achieve desired objectives. NASA has undertaken the task of organizing and leveraging the vast expanse of information in the Heliophysics domain. To achieve this goal NASA has supported the SPASE modeling effort and established domain specific Virtual Observatories. We explore how this effort fits into an information engineering approach and discuss the importance and relevance of data modeling, standardization and open frameworks to achieving NASA's goals. The SPASE modeling effort and independent analysis by some of the Virtual Observatories have determined that the best solution is the development of data systems that are organized based on available resources, that is, adaptive to their environment. We explore why adaptive data systems are the best choice for reducing information entropy, adding value and improving information dissemination within NASA's Heliophysics domain.



Links and References

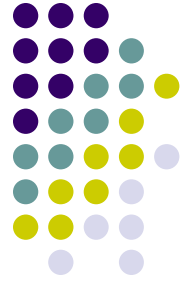
<http://www.ganttthead.com/process/processMain.cfm?ID=2-12009-2>

http://en.wikipedia.org/wiki/Information_engineering



Bibliography

1. Martin, James and Clive Finkelstein. Nov 1981. "Information Engineering", Technical Report (2 volumes), Savant Institute, Carnforth, Lancs, UK.
2. Macdonald, Ian. 1986. "Information Engineering" in "Information Systems Design Methodologies", ed TW Olle et al, North-Holland.
3. Macdonald, Ian. 1988. "Automating the Information Engineering Methodology with the Information Engineering Facility", in "Computerized Assistance during the Information Systems Life Cycle", ed TW Olle et al, North-Holland.
4. Finkelstein, Clive. 1989. "An Introduction to Information Engineering : From Strategic Planning to Information Systems". Sydney: Addison-Wesley.
5. Martin, James. 1989. "Information Engineering" (3 volumes), Prentice-Hall Inc.
6. Hares, John. 1992. "Information Engineering for the Advanced Practitioner", Wiley.



Information Engineering

It encompasses:

- Goals and Objectives
 - Use cases
- Data Engineering
 - Data Models
 - Dictionaries
 - Taxonomies
 - Ontologies
- System Engineering
 - Services
 - Requirements